dbGaP Study Release Notes

Release Notes for NHLBI TOPMed WGS WGHS, phs001040.v5.p1
"NHLBI TOPMed: Novel Risk Factors for the Development of Atrial Fibrillation in Women"

For any questions or comments, please contact: dbgap-help@ncbi.nlm.nih.gov.

October	18, 2016	Version 1 Data set release date
December	13, 2017	Version 2 Data set release date
July	19, 2018	Version 3 Data set release date
April	21, 2020	Version 4 Data set release date
June	2, 2021	Version 5 Data set release date

2021-06-02
Version 5 Data set release for NHLBI TOPMed WGS WGHS now available

This release includes the addition of Freeze 9 whole genome sequences (WGS) and corresponding VCFs. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Health/Medical/Biomedical (HMB)

Data Type	subjects	samples
Phenotype	118	117
Seq_DNA_SNP_CNV (VCFs)	117	117
WGS*	117	117

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.
For a description of non-SRA SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Study and Phenotype Data Updates

1. **New Study Accession**

NHLBI TOPMed WGS WGHS version 4 phs001040.v4.p1 has been updated to version 5. The dbGaP accession for the current set of data is **phs001040.v5.p1**. The participant number (p#) has not changed in version 5. No new subjects have been added to this study.

2. There are no updates to the phenotype datasets.

Molecular Data Updates

Two genomic accessions, phg001377.v1 freeze 8 and phg001570.v1 freeze9, are associated with the study.
1. See download components 'sample-info' for manifest of genotyped samples and files.
2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3. Quality control data are in 'genotype-qc' tars.

phg	freeze	sample_cnt	subject_cnt

phg001377.v1	8	117	117
phg001570.v1	9	117	117

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001040/phs001040.v5.p1

2020-04-21

Version 4 Data set release for NHLBI TOPMed WGS WGHS now available

This release includes the addition of Freeze 8 whole genome sequences (WGS) brokered through the Sequence Read Archive (SRA), and VCFs derived from WGS. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Health/Medical/Biomedical (HMB)

Data Type	subjects	samples
Phenotype	118	117
Seq_DNA_SNP_CNV (VCFs)	117	117
WGS*	117	117

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.
For a description of non-SRA SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Study and Phenotype Data Updates

1. **New Study Accession**

NHLBI TOPMed WGS WGHS version 3 phs001040.v3.p1 has been updated to version 4. The dbGaP accession for the current set of data is **phs001040.v4.p1**. The participant number (p#) has not changed in version 4. No new subjects have been added to this study. No phenotype datasets have been updated.

Molecular Data Updates

1. New molecular data accession, phg001377.v1 containing genotypes on GRC38 is added to the study:
    a. See download components phg001377.v1.TOPMed_WGS_AF_Women_v4.sample-info.MULTI.tar for manifest of genotyped samples and files.
    b. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
    c. Quality control data are in phg001377.v1.TOPMed_WGS_AF_Women_v4.genotype-qc.MULTI.tar.
2. Only Freeze 5b and Freeze 8 VCFs will be available for download.

phg_name	dataset_name	markerset	consent_code	sample_cnt	subject_cnt
phg001065.v1*	TOPMed_WGS_AF_Women_v3	WGS_markerset_grc38	1	98	98
phg001377.v1	TOPMed_WGS_AF_Women_v4	WGS_markerset_grc38	1	117	117

* rollover phg from v3

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001040/phs001040.v4.p1

2018-07-19

Version 3 Data set release for NHLBI TOPMed WGS WGHS now available

This release includes a second genotype call set (GRCh38) and updated phenotype tables. Please refer to the latest study configuration report for a detailed description of each download component.

Changes have been made to the name of the consent group and DUL.

Consent group 1 (c1) is changed from Health/Medical/Biomedical (IRB) (HMB-IRB) to Health/Medical/Biomedical (HMB).

The following is the current Data Use Limitation:

Use of this data is limited to health/medical/biomedical purposes, does not include the study of population origins or ancestry.
Use of the Study data deposited in dbGaP is restricted to research on associations between phenotypes and genotypes.

dbGaP Study Release Notes

For further questions, comments, or concerns, please contact the NHLBI Data Access Committee nhlbigeneticdata@mail.nih.gov.

Consent group 1 (c1): Health/Medical/Biomedical (HMB)

	Phenotype	VCFs	WGS
subjects	118	115	115
samples	117	115	115

Study and Phenotype Data Updates

1. **New Study Accession**

NHLBI TOPMed WGS WGHS version 2 phs001040.v2.p1 has been updated to version 3. The dbGaP accession for the current set of data is **phs001040.v3.p1**. The participant number (p#) has not changed in version 3. No new subjects have been added to this study.

2. **Updated Datasets (n=2 datasets)**

pht	version	Dataset Name
5204	3	TOPMed_WGS_WGHS_Sample
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes

3. **New Variables (n=33 variables)**

pht	pht version	Dataset Name	phv	Variable Name
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375245	APOA1
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375246	APOB
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375247	BMI
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375248	CREA
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375249	CRP
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375250	DBPCAT
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375251	ALCUSE
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375252	FIBRINOGEN
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375253	HBA1C
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375254	HEIGHT
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375255	HCYS
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375256	ICAM
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375257	LPA
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375258	EXERCISE
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375259	SBPCAT
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375260	SMK3CAT
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375261	WEIGHT
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375262	diabetbs
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375263	metabsyn
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375264	agemenarche

5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375265	uterinefibr
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375266	E1
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375267	E2
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375268	E3
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375269	E4
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375270	E5
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375271	E6
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375272	E7
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375273	E8
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375274	E9
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375275	E10
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375276	agemeno
5682	3	TOPMed_WGS_WGHS_Subject_Phenotypes	375277	WHratio

4. Please note we are discontinuing the submission and distribution of the SAMPLE_USE variable. The sample use counts will be populated by SRA (sequences) and dbGaP (all other submitted molecular data).

Molecular Data Updates

New molecular data accession, phg001065.v1 containing genotypes on GRC38 is added to the study:

1. See download components 'sample-info' for manifest of genotyped samples and files included in the release.
2. Genotype calls are available in VCF-formatted multi-sample genotype matrices divided by chromosome ('genotype-calls-vcf').

phg_acc	dataset	data_type	markerset	sample_cnt	subject_cnt
phg000986.v1	TOPMed_WGS_AF_Women_v2	Seq_DNA_SNP_CNV	WGS_markerset_grc37	115	115
phg001065.v1	TOPMed_WGS_AF_Women_v3	Seq_DNA_SNP_CNV	WGS_markerset_grc38	98	98

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001040/phs001040.v3.p1

dbGaP Study Release Notes

2017-12-13

Version 2 Data set release for NHLBI TOPMed WGS WGHS now available

This release includes updated phenotype tables, whole genome sequences (WGS) brokered through the SRA, and VCFs derived from WGS. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Health/Medical/Biomedical (IRB) (HMB-IRB)

	Phenotype	Seq_DNA_SNP_CNV	Seq_DNA_WholeGenome
subjects	118	115	117
samples	117	115	117

Molecular data descriptions:
(https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi)

a. Seq_DNA_SNP_CNV: SNP and CNV genotypes derived from sequence data (VCFs)
b. Seq_DNA_WholeGenome: Whole exome sequencing

Study and Phenotype Data Updates

1. **New Study Accession**

NHLBI TOPMed WGS WGHS version 1 phs001040.v1.p1 has been updated to Version 2. The dbGaP accession for the current set of data is **phs001040.v2.p1**. The participant number (p#) has not changed in version 2. No new subjects have been added to this study.

2. **Updated Datasets (n=4 datasets; all existing variables have been updated)**

pht	version	Dataset Name
5203	2	TOPMed_WGS_WGHS_Subject
5204	2	TOPMed_WGS_WGHS_Sample
5205	2	TOPMed_WGS_WGHS_Sample_Attributes
5682	2	TOPMed_WGS_WGHS_Subject_Phenotypes

3. **New Variables (n=14 variables)**

pht	pht version	Dataset Name	phv	Variable Name
5203	2	TOPMed_WGS_WGHS_Subject	328554	SUBJECT_SOURCE
5203	2	TOPMed_WGS_WGHS_Subject	328555	SOURCE_SUBJECT_ID
5205	2	TOPMed_WGS_WGHS_Sample_Attributes	328563	SEQUENCING_CENTER
5205	2	TOPMed_WGS_WGHS_Sample_Attributes	328564	Funding_Source
5205	2	TOPMed_WGS_WGHS_Sample_Attributes	328565	TOPMed_Phase
5205	2	TOPMed_WGS_WGHS_Sample_Attributes	328566	TOPMed_Project
5205	2	TOPMed_WGS_WGHS_Sample_Attributes	328567	Study_Name
5682	2	TOPMed_WGS_WGHS_Subject_Phenotypes	328556	race
5682	2	TOPMed_WGS_WGHS_Subject_Phenotypes	328557	LDL
5682	2	TOPMed_WGS_WGHS_Subject_Phenotypes	328558	HDL
5682	2	TOPMed_WGS_WGHS_Subject_Phenotypes	328559	TGGB

5682	2	TOPMed_WGS_WGHS_Subject_Phenotypes	328560	CHOL
5682	2	TOPMed_WGS_WGHS_Subject_Phenotypes	328561	FAST8HRS
5682	2	TOPMed_WGS_WGHS_Subject_Phenotypes	328562	cholmeds1

Molecular Data Updates

1. See download components phg000986.v1.TOPMed_WGS_AF_Women_v2.sample-info.MULTI.tar and phg000986.v1.TOPMed_WGS_AF_Women_v2.marker-info.MULTI.tar for manifest of genotyped samples and files, and information about marker set used for genotyping.
2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3. Quality control data are in phg000986.v1.TOPMed_WGS_AF_Women_v2.genotype-qc.MULTI.tar.

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001040/phs001040.v2.p1

2016-10-18

Version 1 Data set release for NHLBI TOPMed WGS WGHS now available

This release includes TOPMed Phase I phenotype tables, whole genome sequences (WGS) brokered through the SRA, and VCFs derived from WGS. Additionally, phenotype tables include subjects and samples beyond TOPMed Phase I in order to instantiate IDs for future versions. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Health/Medical/Biomedical (IRB) (HMB-IRB)

	phenotype	SRA/VCFs
subjects	118	111
samples	118	111

Molecular Data Updates

1. See download components phg000799.v1.TOPMed_WGS_AF_Women.sample-info.MULTI.tar for manifest of genotyped samples.

2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3. Quality control data are in phg000799.v1.TOPMed_WGS_AF_Women.genotype-qc.MULTI.tar.

Authorized Access (Individual Level Data and SRA Data)

Individual level data and SRA sequencing data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001040/phs001040.v1.p1